Panel VAR Models with Spatial Dependence[*]

Jan Mutl
Graduate Assistant, University of Maryland College Park

March 20, 2002

Abstract

I consider a panel vector autoregressive (panel VAR) model with cross-sectional dependence of the model disturbances that can be characterized by a …rst order spatial autoregressive process. I derive asymptotic properties of a constrained maximum likelihood estimator that uses a consistent estimate of the degree of the spatial autocorrelation to concentrate the likelihood function. The asymptotic properties are derived taking the time dimension of the panel as …xed and letting the cross-sectional dimension tend to in…nity.

1. Introduction

Vector autoregressive (VAR) models are extensively used in econometric applications in a wide variety of …elds. The extension to panel data represents an interesting challenge due to the likely presence of cross-sectional heterogeneity. I consider a panel VAR model with …xed time dimension T and derive asymptotic properties of a proposed estimation procedure with respect to the cross-sectional dimension N. When the cross-sectional dimension is …xed, one has to parsimoniously parameterize the correlations across cross-sectional units in order to avoid the incidental parameters problem. In this paper I follow the spatial econometrics literature and study a …rst order spatial autocorrelation model with a known

[*]I am grateful to Michael Binder, Harry Kelejian and Ingmar Prucha for helpful advice and comments. However, all errors and omissions are my responsibility. Correspondence address: Department of Economics, University of Maryland, Tydings Hall, College Park, MD 20742, Email: mutl@econ.bsos.umd.edu.

spatial weighting matrix. The panel spatial autocorrelation model is a generalization of the single cross-section models that include the single equation models, e.g., Cliff and Ord (1973, 1981), and the simultaneous equation models, such as Whittle (1954), Anselin (1988) or Kelejian and Prucha (1998, 1999 and 2001).

On the other hand, the model extends the panel VAR literature to allow for cross-sectional dependence of the model disturbances; for models with homogeneous disturbances see, e.g., Binder et al. (2001) for the quasi maximum likelihood (QML) and minimum distance (MD) estimators or Arellano and Bond (1991), Ahn and Schmidt (1995) or Arellano and Bover (1995) for the generalized method of moments (GMM) approach.

Existing extensions of panel models for cross-sectional dependence of the model disturbances include a generalized least squares test to test for unit roots in a panel data (although without deriving any asymptotic properties of the estimator) in O'Connell (1998), a two-step sieve least squares procedure to estimate a panel VAR model with a nondiagonal cross-sectional covariance matrix that is proportional to an observed economic distance measure in Chen and Conley (2000) who look at asymptotics when the cross-sectional dimension is fixed, and, finally, Chang (2001) who derives asymptotic properties of a univariate panel model with a general unrestricted form of cross-sectional heterogeneity when the cross-sectional dimension of the panel is also fixed. This approach is complementary to the present paper which considers asymptotics with respect to the cross-sectional dimension and keeping the time dimension fixed.

2. The Panel VAR Model

In this section I specify the model and discuss the main assumptions that will be maintained in the consistency proofs. The specification adopted here follows the spatial autoregressive framework with known spatial weighting matrix. In such models the correlation across agents is conveniently parameterized with only one parameter. The the model can be expressed as

$$\mathbf{y}_{it} = (\mathbf{I}_m - \mathbf{\Phi})\mu_i + \mathbf{\Phi}\mathbf{y}_{i,t-1} + \mathbf{u}_{it} \tag{1}$$

$$\mathbf{u}_{it} = \rho \sum_{j=1}^{N} \mathbf{w}_{ij,t}\mathbf{u}_{jt} + \varepsilon_{it} \tag{2}$$

where the first subscript $i \in \{1, \ldots, N\}$ refers to the cross-sectional dimension and the second subscript $t \in \{1, \ldots, T\}$ refers to the time dimension of the panel of observations $\{\mathbf{y}_{it}\}_{1 \leq t \leq T}^{1 \leq i \leq N}$. I also allow the model to contain more than one equation and so the observations $\mathbf{y}_{it}$, the individual-specific effects μ_i and the disturbances $\mathbf{u}_{it}$ and ε_{it} are $m \times 1$ vectors and the known weighting parameters $\mathbf{w}_{ij}$, the unknown model parameters $\mathbf{\Phi}$ and the identity matrix $\mathbf{I}_m$ are $m \times m$ matrices. The degree of spatial autocorrelation is captured by the scalar parameter λ.

Stacking across individuals we obtain

$$\mathbf{y}_t = (\mathbf{I}_N \otimes [\mathbf{I}_m - \mathbf{\Phi}])\mu + (\mathbf{I}_N \otimes \mathbf{\Phi})\mathbf{y}_{t-1} + \mathbf{u}_t \tag{3}$$

$$\mathbf{u}_t = \lambda \mathbf{W}_t \mathbf{u}_t + \varepsilon_t \tag{4}$$

where

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_{1t} \\ \vdots \\ \mathbf{y}_{Nt} \end{pmatrix}_{mN \times 1} \qquad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}_{mN \times 1} \tag{5}$$

$$\mathbf{u}_t = \begin{pmatrix} \mathbf{u}_{1t} \\ \vdots \\ \mathbf{u}_{Nt} \end{pmatrix}_{mN \times 1} \qquad \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{pmatrix}_{mN \times 1}$$

and the $mN \times mN$ weighting matrix $\mathbf{W}_t$ is

$$\mathbf{W}_t = \begin{pmatrix} \mathbf{w}_{11;t} & \cdots & \mathbf{w}_{1N;t} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{N1;t} & \cdots & \mathbf{w}_{NN;t} \end{pmatrix}_{mN \times mN} \tag{6}$$

Solving for the disturbance terms yields

$$\mathbf{u}_t = (\mathbf{I}_{mN} - \lambda \mathbf{W}_t)^{-1} \varepsilon_t \tag{7}$$

To facilitate identification of the model, I assume that there is no spatial correlation across equations, that is each $m \times m$ matrix $\mathbf{w}_{ij}$ is diagonal. However, the model allows for contemporaneous correlation across equations in different cross-sections because the variance-covariance matrix of the error terms ε_{it} is left unrestricted.[1]

[1] There is cross-equation correlation for a single cross-section and since the cross-sections are spatially correlated, the error terms in different equations for different cross-sections will be contemporaneously correlated.

2.1. Random vs. Fixed Effects Specification

Allowing for individual effects without any additional restrictions, such as random or fixed effects specification, leads to an incidental parameters problem. As the time dimension of the panel is fixed, one cannot consistently estimate a general form of the individual-specific effects with a finite number of observations per parameter. To resolve this problem, there are two options. Either to assume that there is a well-behaved distribution (e.g. with finite fourth moments) from which the individual-specific effects are generated (the random effects specification), or transform the data to obtain specification that does not contain the individual-specific effects (the fixed effect specification). The usual approach in the fixed effect specification is to first-difference the data; see the argument in Hsiao, Pesaran and Tahmiscioglu (2001) who show in a univariate context that the QML estimator is invariant to the choice of the transformation matrix that eliminates the individual-specific effects. The argument is readily extended to the multivariate setting in this paper. However, the fixed effect specification and first-differencing does not eliminate the incidental parameter problem unless we assume that the spatial weighting matrices are constant over time. Hence the choice between fixed and random effects specification depends on which of the two assumptions (constant weighting matrix or existence of the distribution that generates the individual-specific effects) is more appropriate.

In this paper I work out the case of fixed effects with constant spatial weighting matrix. However, the extension to random effects with time varying spatial weighting matrix is straightforward.

2.2. Initial Disturbances Specification

Instead of conditioning on initial observations, I explicitly treat the initial conditions when defining the likelihood function. There are several assumptions one can make. The least restrictive case is worked out in this paper. Denote the vector of initial model disturbances as

$$\mathbf{u}_0 = \mathbf{y}_0 - \mu \tag{8}$$

I assume that $\mathbf{u}_0$ is spatially correlated and is generated by

$$\mathbf{u}_0 = \lambda \mathbf{W}\mathbf{u}_0 + \xi \tag{9}$$

where ξ is an $N \times 1$ vector of independently and identically distributed (in N) initial random disturbances.

Hence the initial observations are

$$\begin{aligned}
\Delta \mathbf{y}_1 &= \mathbf{u}_1 - [\mathbf{I}_N \otimes (\mathbf{I}_m - \mathbf{\Phi})]\mathbf{u}_0 \\
&= (\mathbf{I}_N - \rho\mathbf{W})^{-1}(\varepsilon_1 - [\mathbf{I}_N \otimes (\mathbf{I}_m - \mathbf{\Phi})]\eta)
\end{aligned} \tag{10}$$

Notice that this implies that

$$(\varepsilon_1 - [\mathbf{I}_N \otimes (\mathbf{I}_m - \mathbf{\Phi})]\eta) = (\mathbf{I}_N - \rho\mathbf{W})\Delta \mathbf{y}_1 \tag{11}$$

We assume that the initial disturbances are independent of subsequent error terms. We use the notation

$$\mathrm{var}(\varepsilon_{i1} - (\mathbf{I}_m - \mathbf{\Phi})\eta_i) = \mathbf{\Psi} \quad \text{and} \quad \mathrm{var}(\varepsilon_{it}) = \mathbf{\Omega}_\varepsilon \tag{12}$$

Given that $\mathbf{\Phi} \neq \mathbf{I}_m$, have that

$$\mathbf{\Psi} = \mathbf{\Omega}_\varepsilon + [\mathbf{I}_N \otimes (\mathbf{I}_m - \mathbf{\Phi})]\mathrm{var}(\eta_i)[\mathbf{I}_N \otimes (\mathbf{I}_m - \mathbf{\Phi})] \tag{13}$$

and hence $\mathbf{\Psi}$ is unconstrained.

In general, if the eigenvalues of $\mathbf{\Phi}$ are inside the unit circle, one could make further assumptions on the η disturbances and express $\mathbf{\Psi}$ in terms of $\mathbf{\Phi}$ and $\mathbf{\Omega}_\varepsilon$. In particular, since in this case the data generating process is stationary and, therefore, one could assume that it has started in an infinite past. This implies that the initial observations $\mathbf{y}_0$ are drawn from the limiting stationary distribution of the process, e.g. that:

$$\mathbf{y}_0 = \sum_{j=0}^{\infty} [\mathbf{I}_N \otimes \mathbf{\Phi}]^{j-1}\varepsilon_{-j} \tag{14}$$

Therefore

$$\Delta \mathbf{y}_1 = (\mathbf{I}_N - \rho\mathbf{W})^{-1} \sum_{j=0}^{\infty} [\mathbf{I}_N \otimes \mathbf{\Phi}]^{j-1}\Delta\varepsilon_{-j} \tag{15}$$

and

$$\begin{aligned}
\mathrm{var}(\varepsilon_1 - [\mathbf{I}_N \otimes (\mathbf{I}_m - \mathbf{\Phi})]\eta) &= \mathrm{var}\left([\mathbf{I}_N - \rho\mathbf{W}]\Delta \mathbf{y}_1\right) \\
&= \mathrm{var}\left(\sum_{j=0}^{\infty} [\mathbf{I}_N \otimes \mathbf{\Phi}]^{j-1}\Delta\varepsilon_{-j}\right) \\
&= \mathrm{var}\left(\varepsilon_0 + (\mathbf{I}_N \otimes [\mathbf{I}_m - \mathbf{\Phi}]) \sum_{j=0}^{\infty} [\mathbf{I}_N \otimes \mathbf{\Phi}]^{j}\varepsilon_{-j-1}\right) \\
&= \mathbf{\Omega}_\varepsilon + (\mathbf{I}_N \otimes [\mathbf{I}_m - \mathbf{\Phi}]) \left(\sum_{j=1}^{\infty} \mathbf{\Phi}^{j}(\mathbf{I}_N \otimes \mathbf{\Omega}_\varepsilon)\mathbf{\Phi}'^{j}\right) (\mathbf{I}_N \otimes [\mathbf{I}_m - \mathbf{\Phi}])'
\end{aligned} \tag{16}$$

Such assumption complicates the algebra and we leave this for further extensions of our model. In the following we treat vech$\boldsymbol{\Psi}$ as a vector of additional free parameters.

2.3. Maintained Assumptions

To be able to derive the asymptotic properties of the model I make the following assumptions about the disturbances and the spatial weighting matrices.

Assumption 1. The disturbance vector ε_{it} is identically and independently distributed with zero mean, finite positive-definite variance matrix $\boldsymbol{\Omega}_{\varepsilon}$ and finite fourth moments.

The above assumption is needed to ensure that the observable data, which is a transformation of the ε_{it} process, has a well-defined asymptotic properties.

The following assumption is necessary for identification of the model:

Assumption 2. The diagonal elements of each $\mathbf{W}_t$ are zero and each $\mathbf{w}_{it}$ matrix is diagonal.

The next two assumptions ensure that the weighting matrices do not 'explode' as the sample size increases.

Assumption 3. The matrices $(\mathbf{I}_{mN} - \lambda \mathbf{W}_t)$ are nonsingular for all $|\lambda| < 1$.

Assumption 4. The row and column sums of the matrices $\mathbf{W}_t$ and $(I_N - \lambda \mathbf{W}_t)^{-1}$ are bounded uniformly in absolute value.

3. Estimation

The model can be estimated using a variety of approaches. Straightforward least squares estimation of the first differences of the observations on its lagged values is not consistent because the error term $\Delta \mathbf{u}_t$ is correlated with the explanatory variable $\Delta \mathbf{y}_{t-1}$. However, I show that an instrumental variable (IV) estimation leads to a consistent estimates of the spatially correlated disturbances. We can then use a method of moments estimation and obtain a consistent estimator of the spatial parameter λ; e.g. use the moment conditions based on the estimated disturbances:

$$\Delta \hat{\mathbf{u}}_t = \Delta \mathbf{y}_t - (\mathbf{I}_N \otimes \hat{\boldsymbol{\Phi}}_{IV}) \Delta \mathbf{y}_{t-1} \tag{17}$$

where $\hat{\boldsymbol{\Phi}}_{IV}$ is the IV estimators of $\boldsymbol{\Phi}$. I show that this two stage procedure leads to a consistent estimator of λ. Kelejian and Prucha (1999) show consistency of a similar two stage procedure for model with spatial lags in both the dependent variable as well as the error term.

Finally, we can use the spatial Cochrane-Orcutt transformation and write the model as

$$(\mathbf{I}_{mN} - \lambda\mathbf{W})\Delta\mathbf{y}_t = (\mathbf{I}_{mN} - \lambda\mathbf{W})(\mathbf{I}_N \otimes \mathbf{\Phi})\Delta\mathbf{y}_{t-1} + \Delta\varepsilon_t \tag{18}$$

If λ is known, the transformed model can be estimated with standard techniques, such as the QML method in Binder, Hsiao, Mutl and Pesaran (2002) or the GMM approach as in Arellano and Bond (1991), Ahn and Schmidt (1995) or Arellano and Bover (1995). However, since λ has to be estimated, we need to prove that it is a nuisance parameter.

In the following, I first define the IV estimator and show that it produces consistent estimates of the disturbance terms. I then discuss the moments estimator of the spatial parameter. Finally, I define the full as well as the constrained QML procedures and show that the spatial parameter is a nuisance parameter.

3.1. Instrumental Variable Estimation

To be able to define the IV estimator, it turns out to be convenient to stack the model differently. Our model is:

$$\Delta\mathbf{y}_{it} = \mathbf{\Phi}\Delta\mathbf{y}_{i,t-1} + \Delta\mathbf{u}_{it} \tag{19}$$

where $\Delta\mathbf{y}_{it}$ and $\Delta\mathbf{u}_{it}$ are $m \times 1$ vectors. After taking transpose and staking the observations at different times for a given cross-section, we have

$$\begin{pmatrix} \Delta\mathbf{y}'_{i1} \\ \vdots \\ \Delta\mathbf{y}'_{iT} \end{pmatrix}_{T\times m} = \begin{pmatrix} \Delta\mathbf{y}'_{i0} \\ \vdots \\ \Delta\mathbf{y}'_{i,T-1} \end{pmatrix}_{T\times m} \mathbf{\Phi}'_{m\times m} + \begin{pmatrix} \Delta\mathbf{u}'_{i1} \\ \vdots \\ \Delta\mathbf{u}'_{iT} \end{pmatrix}_{T\times m} \tag{20}$$

or with the obvious notation

$$\Delta\mathbf{Y}_i = \Delta\mathbf{Y}_{i,-1}\mathbf{\Phi}' + \Delta\mathbf{U}_i \tag{21}$$

Stacking the cross-sections yields

$$\Delta\mathbf{Y} = \Delta\mathbf{Y}_{-1}\mathbf{\Phi}' + \Delta\mathbf{U} \tag{22}$$

where $\Delta\mathbf{Y} = (\Delta\mathbf{Y}'_1; \ldots; \Delta\mathbf{Y}'_N)'$, $\Delta\mathbf{Y}_{-1} = (\Delta\mathbf{Y}'_{1,-1}; \ldots; \Delta\mathbf{Y}'_{N,-1})'$ and $\Delta\mathbf{U} = (\Delta\mathbf{U}'_1; \ldots; \Delta\mathbf{U}'_N)'$.

We define the IV estimator of $\mathbf{\Phi}$ as

$$\hat{\mathbf{\Phi}}_{IV} = \left[\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right]^{-1}\hat{\mathbf{Z}}'\Delta\mathbf{Y} \tag{23}$$

where $\hat{\mathbf{Z}} = \mathbf{P}_H \Delta \mathbf{Y}$ with $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$ where $\mathbf{H}$ is vector of instruments used for $\Delta \mathbf{Y}_{-1}$. I suggest the use of the instruments $\mathbf{H} = \mathbf{Y}_{-2} = (\mathbf{Y}'_{1;-2}; ::::; \mathbf{Y}'_{N;-2})'$ where $\mathbf{Y}_{i;-2} = (\mathbf{y}_{i;-1}; ::::; \mathbf{y}_{i;T-2})'$. However, any instruments that satisfy the following conditions, lead to consistent estimates of the spatially correlated disturbances:

Assumption 5. The instrument matrix $\mathbf{H}$ has a full column rank.

Assumption 6. The instruments satisfy the following:

1. $\text{plim}\, \frac{1}{N}\mathbf{H}'\mathbf{H} = \mathbf{Q}_{HH}$ where $\mathbf{Q}_{HH}$ is finite and nonsingular;
2. $\text{plim}\, \frac{1}{N}\mathbf{H}'\Delta\mathbf{Y} = \mathbf{Q}_{HY}$ where $\mathbf{Q}_{HY}$ is finite and has a full column rank.
3. The instruments $\mathbf{H}$ can be expressed as $\mathbf{H} = \mathbf{F}(\xi_1; :::; \xi_m)$ where each ξ_i is a $NT \times 1$ vector of identicaly and independently distributed random variables and $\mathbf{F}$ is an $NT \times NT$ nonstochastic absolutely summable matrix. Furthermore, each ξ_i is independent of ε_{it}.

The first two assumptions guarantee that the instruments are not degenerate and that they are asymptoticaly correlated with the variables they replace. The last assumption implies that the instruments are not correlated with the error terms and that our central limit theorem can be applied. Given these additional assumptions we can assert that the IV estimation produces consistent estimates:

Proposition 1. Given the setup and assumptions 1-6, the IV estimator is consistent and the rate of convergence is $N^{-1/2}$; that is $\hat{\Phi}_{IV} = \Phi + O_p(N^{-1/2})$.

Remark: The rate of convergence is important for consistency of estimation λ (the degree of spatial correlation in the residuals) in the the second step of the procedure.

Proof of proposition 1.

Substituting for the instruments in the model yields

$$\begin{aligned} \hat{\Phi}_{IV} &= \Phi + \left[\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right]^{-1}\hat{\mathbf{Z}}'\Delta\mathbf{U} \qquad (24) \\ &= \Phi + \left[\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right]^{-1}\Delta\mathbf{Y}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\Delta\mathbf{U} \end{aligned}$$

To show consistency we prove that

$$\text{plim}\, \frac{1}{N}\hat{\mathbf{Z}}'\hat{\mathbf{Z}} = \mathbf{Q}_{ZZ} \qquad (25)$$

where $\mathbf{Q}_{ZZ}$, is finite and nonsingular and provide a central limit theorem (CLM) for the remaining $N^{-1/2}\mathbf{H}'\cdot\mathbf{U}$ term. Together, these results imply that

$$N^{1/2}(\hat{\mathbf{\Phi}}_{IV} - \mathbf{\Phi}) \to N(0;\mathbf{Q}) \tag{26}$$

in distribution; where

$$\mathbf{Q} = \mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{HY}'\mathbf{Q}_{HY}\mathbf{Q}_{ZZ}^{-1} \ \ p\lim_{N\to\infty} tr(\cdot\mathbf{U}\cdot\mathbf{U}') \tag{27}$$

The existence and nonsingularity of the $\mathbf{Q}_{xx}$ matrices is due to the correct choice of instruments. The CLM utilized here is a modification of the CLM for quadratic forms of triangular arrays given, for example, in Kelejian and Prucha (2001) or in Pinkse (1999).

Theorem 2. Central Limit Theorem (CLM)

Let ς_1 and ς_2 be two vectors each consisting of n independent and identically distributed zero mean random disturbances (with finite 4th moments), furthermore let ς_1 and ς_2 be independent of each other (hence expected value of the quadratic form is zero). Let A_n be an $n\times n$ nonstochastic absolutely summable matrix. Then

$$\frac{\varsigma_1'\mathbf{A}_n\varsigma_2}{var(\varsigma_1'\mathbf{A}_n\varsigma_2)} \to N(0;1) \tag{28}$$

in distribution.

To be able to apply the CLM we first express the instrument as in assumption 6 and then apply the above CLM to each element of $\mathbf{H}'\cdot\mathbf{U}$ separately. We have

$$\begin{aligned}\mathbf{H}'\cdot\mathbf{U} &= (\varsigma_1;\ldots;\varsigma_m)'\mathbf{F}'\cdot\mathbf{U} \\ &= (\varsigma_1;\ldots;\varsigma_m)'\mathbf{F}'(\cdot\mathbf{U}_1';\ldots;\cdot\mathbf{U}_N')'\end{aligned} \tag{29}$$

Hence $ij-$th element of $\mathbf{H}'\cdot\mathbf{U}$ is of the form

$$\varsigma_i'\mathbf{F}'\begin{pmatrix}\cdot\mathbf{U}_1\\ \vdots \\ \cdot\mathbf{U}_N\end{pmatrix}_{.,j} = \varsigma_i'\mathbf{F}'\mathbf{T}(\mathbf{I}_{mN} - \lambda\mathbf{W})\varepsilon_t \tag{30}$$

where $\mathbf{T}$ is a finite and absolutely summable tranformation matrix whose elements depend of only on N. Since elements of ε_{it} are independent of the elements ς_j, the conditions of our CLM are met and $N^{-1/2}\mathbf{H}'\Delta\mathbf{U}$ covnverges in distribution. QED.

Note that our suggested instruments meet the required conditions. By backward substitution we can eliminate the lagged dependent variables and express the instruments as a function of lagged disturbance terms and lagged explanatory variables. It is easily verified that

$$\Delta\mathbf{y}_t = (\mathbf{I}_{mN} - \lambda\mathbf{W})^{-1}\left(\sum_{j=0}^{t-2}(\mathbf{I}_N \otimes \mathbf{\Phi})^{j-1}\Delta\varepsilon_{t-j} + (\mathbf{I}_N \otimes \mathbf{\Phi})^{t-1}[\varepsilon_1 - (\mathbf{I}-\mathbf{\Phi})\mu]\right) \tag{31}$$

and hence we have that

$$\mathbf{H} = (\mathbf{Y}'_{1,-2}, \ldots, \mathbf{Y}'_{N,-2})' \tag{32}$$

$$= \mathbf{F}(\varepsilon_1 - (\mathbf{I}-\mathbf{\Phi})\mu, \Delta\varepsilon_2, \ldots, \Delta\varepsilon_{T-2})' \tag{33}$$

where

$$\mathbf{F} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda\mathbf{W}')]^{-1}\left[\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda\mathbf{W})^{-1}\right]$$

Our assumptions on the spatial weighting matrices imply that the $N \times N$ matrix $\mathbf{F}$ is absolutely summable.

3.2. Estimation of λ

The second step in the proposed estimation procedure is to use moment conditions based on the estimated disturbances:

$$\Delta\hat{\mathbf{u}}_t = \Delta\mathbf{y}_t - (\mathbf{I}_N \otimes \hat{\mathbf{\Phi}}_{IV})\Delta\mathbf{y}_{t-1} \tag{35}$$

where $\hat{\mathbf{\Phi}}_{IV}$ is the IV estimators of $\mathbf{\Phi}$. Kelejian and Prucha (1999) show consistency of a similar two stage procedure for model with spatial lags in both the dependent variable as well as the error term. The conditions of their theorem 2 are met in the present setup and hence their moment estimator produces a consistnet estimate of the spatial parameter λ. In the appendix, show that λ is a nuisance parameter in a model with $m = 1$. Generalization to $m > 1$ is involves somewhat tedious notation and is omitted here. To demonstrate that λ is nuisance, I show that the off-diagonal elements of the Hessian of the likelihood function, corresponding to the parameter λ, are $o_p(1)$.

3.3. Quasi Maximum Likelihood (QML) Estimation

The likelihood function for the panel VAR model is easily derived under the assumption that $\varepsilon_{it} \sim N(0; \boldsymbol{\Omega}_\varepsilon)$ where $\boldsymbol{\Omega}_\varepsilon$ is the $m \times m$ variance-covariance matrix of ε_{ti}. I specify the exact distribution of the initial observations as in Binder et al. (2001) and derive the QML function taking this into account. We can define the $mN(T+1) \times 1$ vector

$$\Delta\xi = \begin{pmatrix} \Delta \mathbf{y}_0 \\ \Delta \varepsilon_1 \\ \vdots \\ \Delta \varepsilon_T \end{pmatrix} \tag{36}$$

We then have that $E(\Delta\xi) = 0$ and $var(\Delta\xi) = \Sigma_{\Delta\xi}$ where

$$\Sigma_{\Delta\xi} = \begin{pmatrix} \boldsymbol{\Psi} & -\boldsymbol{\Omega}_\varepsilon & & 0 \\ -\boldsymbol{\Omega}_\varepsilon & 2\boldsymbol{\Omega}_\varepsilon & & \\ & & \ddots & -\boldsymbol{\Omega}_\varepsilon \\ 0 & & -\boldsymbol{\Omega}_\varepsilon & 2\boldsymbol{\Omega}_\varepsilon \end{pmatrix} \otimes \mathbf{I}_N \tag{37}$$

with $\boldsymbol{\Psi}$ being a $m \times m$ symmetric matrix of parameters. This specification leaves the variance-covariance matrix of the initial observations unrestricted - e.g. there are $m(m+1)/2$ free parameters.

The likelihood function for the entire sample is then

$$\begin{aligned} L_N(\theta) &= \text{const} - \frac{N}{2}\ln|\boldsymbol{\Sigma}_{\Delta\xi}| + \ln\left|\mathbf{I}_{mN(T+1)} - \lambda\bar{\mathbf{W}}\right| \\ &\quad - \frac{N}{2}\text{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \lambda\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\Delta\xi}^{-1}(\mathbf{I}_{mN(T+1)} - \lambda\bar{\mathbf{W}}')\mathbf{R}\mathbf{S}_N\right] \end{aligned} \tag{38}$$

where $\theta = (\text{vech}\boldsymbol{\Psi}', \text{vech}\boldsymbol{\Omega}_\varepsilon', \text{vec}\boldsymbol{\Phi}')$ is the vector of parameters. The $mN(T+1) \times mN(T+1)$ observable time-invariant spatial weighting matrix $\bar{\mathbf{W}}$ is

$$\bar{\mathbf{W}} = \begin{pmatrix} \mathbf{I}_{mN} & 0 & \cdots & 0 \\ 0 & \mathbf{W}_1 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \mathbf{W}_T \end{pmatrix} \tag{39}$$

The $mN(T+1) \times mN(T+1)$ matrix $\mathbf{R}$ is defined as

$$\mathbf{R} = \mathbf{I}_N \otimes \begin{pmatrix} \mathbf{I}_m & & & 0 \\ -\mathbf{\Phi} & \mathbf{I}_m & & \\ & & \ddots & \\ 0 & & -\mathbf{\Phi} & \mathbf{I}_m \end{pmatrix} \tag{40}$$

and the matrix $\mathbf{S}_N$ is

$$\mathbf{S}_N = \mathbf{I}_N \otimes \frac{1}{N} \sum_{i=0}^{N} \Delta \mathbf{y}_i \Delta \mathbf{y}_i' \tag{41}$$

with $\Delta \mathbf{y}_i = (\Delta \mathbf{y}_{i0}', \ldots, \Delta \mathbf{y}_{iT}')'$ being the vector of the first differences of the observations for the i-th cross-section.

3.3.1. Computational Issues

The computation of the likelihood function should exploit the structure of the $[\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda \mathbf{W})]$ and $\mathbf{\Sigma}_{\Delta\eta}$ matrices when evaluating their determinants and inverses. In particular, we can express $\mathbf{\Sigma}_{\Delta\eta}$ as

$$\mathbf{\Sigma}_{\Delta\eta} = \begin{pmatrix} \mathbf{\Psi} & (\mathbf{A}_1 \otimes \mathbf{\Omega}_\varepsilon) \\ (\mathbf{A}_1' \otimes \mathbf{\Omega}_\varepsilon) & (\mathbf{A}_2 \otimes \mathbf{\Omega}_\varepsilon) \end{pmatrix} \tag{42}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are matrices of constants. The inverse of $\mathbf{\Sigma}_{\Delta\eta}$ is then

$$\mathbf{\Sigma}_{\Delta\eta}^{-1} = \begin{pmatrix} \mathbf{D}^{-1} & -\mathbf{D}^{-1}(\mathbf{A}_1\mathbf{A}_2^{-1} \otimes \mathbf{\Omega}_\varepsilon) \\ (\mathbf{A}_2^{-1}\mathbf{A}_1' \otimes \mathbf{\Omega}_\varepsilon)\mathbf{D}^{-1} & \mathbf{D}^{-1} - (\mathbf{A}_2^{-1}\mathbf{A}_1' \otimes \mathbf{\Omega}_\varepsilon)\mathbf{D}^{-1}(\mathbf{A}_1\mathbf{A}_2^{-1} \otimes \mathbf{\Omega}_\varepsilon) \end{pmatrix} \tag{43}$$

where $\mathbf{D} = \mathbf{\Psi} - (\mathbf{A}_1\mathbf{A}_2^{-1}\mathbf{A}_1 \otimes \mathbf{\Omega}_\varepsilon)$.

3.4. Constrained QML Estimation

Although the QML estimation based on the likelihood function (38) is feasible[2], it is extremely difficult to prove its consistency and asymptotic normality. In this paper, I propose an alternative approach that takes a consistent estimator of the

[2] The QML estimator is likely to be computationally expensive due to the necessity to calculate eigenvalues of a sparse matrix $(I - \lambda W_t)$ which is of the dimension N. With large N this becomes a very demanding problem.

spatial correlation parameter λ and maximizes a constrained likelihood function. That is, maximize

$$
\begin{aligned}
Q_N(\tilde{\theta}) = \; & \text{const} - \frac{N}{2}\ln|\boldsymbol{\Sigma}_{\Delta\eta}| + \ln\left|\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}\right| \\
& - \frac{N}{2}\operatorname{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{m(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}\mathbf{S}_N\right]
\end{aligned} \tag{44}
$$

with respect to $\tilde{\theta} = (\text{vech}\boldsymbol{\Psi}'; \text{vech}\boldsymbol{\Omega}_{\upsilon}'; \text{vec}\boldsymbol{\Phi}')'$, taking the consistent estimator $\hat{\lambda}$ of λ as given. The consistent estimator of the spatial correlation be based on the two-step procedure proposed above.

4. Asymptotics

To prove consistency of the constrained QML estimator, I first prove that the likelihood function converges point-wise in probability to some function (which is the limit of the expected likelihood). I then use identification conditions to show consistency.

4.1. Point-wise Convergence of the Likelihood Function

The constrained likelihood function is

$$
\begin{aligned}
Q_N(\tilde{\theta}) = \; & \text{const} - \frac{N}{2}\ln|\boldsymbol{\Sigma}_{\Delta\eta}| + \ln\left|\mathbf{I}_{m(T+1)} - \hat{\lambda}\bar{\mathbf{W}}\right| \\
& - \frac{N}{2}\operatorname{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}\mathbf{S}_N\right]
\end{aligned} \tag{45}
$$

For the first part we need to show that

$$
\sup_{\theta \in \Theta}\left|\frac{1}{N}Q_N(\tilde{\theta}) - E\left(\frac{1}{N}L_N(\lambda;\theta)\right)\right| \to 0 \quad \text{in probability} \tag{46}
$$

where $\tilde{\theta} = (\text{vech}\boldsymbol{\Psi}'; \text{vech}\boldsymbol{\Omega}'; \text{vec}\boldsymbol{\Phi}')'$ is a vector of parameters from admissible parameter space $\boldsymbol{\Theta}$, e.g. $\boldsymbol{\Omega}$ and $\boldsymbol{\Psi}$ are symmetric positive-definite, etc.

Note that

$$
\frac{1}{N}Q_N(\tilde{\theta}) - E\left(\frac{1}{N}L_N(\lambda;\theta)\right) = \tag{47}
$$

$$
\frac{N}{2}\operatorname{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}(E(\mathbf{S}_N) - \mathbf{S}_N)\right]
$$

$$< \frac{N}{2}\mathrm{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\mathbf{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}\right] \cdot \mathrm{tr}(E(\mathbf{S}_N) - \mathbf{S}_N) \quad (48)$$

To complete the proof, we need to show that

(i) $\mathrm{tr}\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\mathbf{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}$ is bounded as $N \to \infty$, and that

(ii) $\mathrm{tr}\mathbf{S}_N = \frac{1}{N}\sum_{i=0}^{N}\Delta\mathbf{y}_i\Delta\mathbf{y}_i'$ converges to its expected value $\mathrm{tr}\mathbf{R}^{-1}(\mathbf{I}_{m(T+1)} - \hat{\lambda}\bar{\mathbf{W}})^{-1}\mathbf{\Sigma}_{\Delta\eta}(\mathbf{I}_{m(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')^{-1}\mathbf{R}'^{-1}$ in probability as $N \to \infty$.

The former follows from the assumptions I have made on the weighting matrices, specifically the absolute summability of the $\mathbf{W}$ matrix. The latter follows from the i.i.d. assumptions on the disturbance term and the restrictions on the weighting matrices - we have that

$$\begin{aligned}
\mathrm{tr}\mathbf{S}_N &= \mathrm{tr}\frac{1}{N}\Delta\mathbf{y}\mathbf{\Delta}\mathbf{y}' && (49)\\
&= \mathrm{tr}\left[\mathbf{R}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})^{-1}\left(\frac{1}{N}\varepsilon\varepsilon'\right)(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')^{-1}\mathbf{R}'^{-1}\right] && (50)\\
&= \mathrm{tr}\left[(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')^{-1}\mathbf{R}'^{-1}\mathbf{R}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})^{-1}\left(\frac{1}{N}\varepsilon\varepsilon'\right)\right] && (51)\\
&< \mathrm{tr}\left[(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')^{-1}\mathbf{R}'^{-1}\mathbf{R}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})^{-1}\right] \cdot \mathrm{tr}\left(\frac{1}{N}\varepsilon\varepsilon'\right) && (52)
\end{aligned}$$

We can show that $\mathrm{tr}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')^{-1}\mathbf{R}'^{-1}\mathbf{R}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})^{-1}$ is also bounded. Given the assumptions on the disturbances imply $\frac{1}{N}\varepsilon\varepsilon' \to \mathbf{\Sigma}_{\Delta\eta}$, we have the result (ii).

4.2. Consistency

Given the point-wise convergence of the constrained likelihood function, any sequence of $\hat{\theta}_N = \arg\max_{\theta\in\Theta}\left(Q_N(\tilde{\theta})\right)$ has to converge to a set of global maxima of the limiting function $Q(.) = p\lim_{N\to\infty} E[L_N(.)]$. The set of global maxima of this function contains the true parameter vector (this is a direct consequence of Jensen's inequality), furthermore, if some identification conditions are met, it only contains the true value of the parameter and we have consistency.

To judge whether the likelihood has a flat top one needs to inspect asymptotic behavior of the Hessian, i.e. inspect the limits of the $\mathbf{H}_{ij}$ matrices (defined in the appendix). If the probability limit as $N \to \infty$ of some of the $\mathbf{H}_{ij}$ matrices is singular, then the corresponding parameter is asymptotically not identified. The terms in these matrices are similar to the ones discussed in the proofs above, e.g.

$\boldsymbol{\Sigma}_{\epsilon}^{\prime -1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}\mathbf{S}_N'\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\epsilon}^{\prime -1}$
$\boldsymbol{\Sigma}_{\epsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}\mathbf{S}_N$, $\boldsymbol{\Sigma}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')$
$(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\epsilon}^{\prime -1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')$ and
$\mathbf{R}'(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}})\boldsymbol{\Sigma}_{\epsilon}^{\prime -1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')$.

Let us take the expression $\boldsymbol{\Sigma}_{\epsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R}\mathbf{S}_N$ as an example. We need to show that

$$\lim_{N \to \infty} N\boldsymbol{\Sigma}_{\epsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')\mathbf{R} > 0 \tag{53}$$

and that

$$\operatorname*{plim}_{N \to \infty} \frac{1}{N}\mathbf{S}_N > 0 \tag{54}$$

The later plim was discussed above. The former follows from the assumptions on the weighting matrices and the parameter space; since $\hat{\lambda}$ is consistent, the matrix $(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\bar{\mathbf{W}}')$ is not singular as N tends to infinity.

Inspection of the Hessians does not rule out multiplicity of peaks of the likelihood function. However, since the likelihood function is smooth, it does imply local identification and hence consistency of the constrained QML estimator. The following proposition summarizes the main result.

Proposition 3. Under Assumptions 1-4, there exists a neighborhood of the true parameter value $\boldsymbol{\Theta}_1 \subseteq \boldsymbol{\Theta}$ such that maximization of the constrained maximum likelihood function over $\boldsymbol{\Theta}_1$ gives a consistent estimate of $\tilde{\theta}$.

The above proposition implies that as long as the admissible parameter space is compact within some neighborhood of the true parameter value and that we can obtain starting values from this compact neighborhood, the constrained maximum likelihood estimator will be consistent.

5. Conclusion

This paper develops an estimation approach for a panel VAR model with spatial dependence. I suggest a three-step estimation procedure. In the first step, instrumental variables procedure is used to consistently estimate the spatially correlated disturbances. In the second step, a method of moments estimation is used to obtain a consistent estimate of the spatial parameter. The final step of the procedure could be either a constrained maximum likelihood procedure or moments estimation based on a model transformed by a spatial Cochrane-Orcutt transformation.

I introduce the constrained maximum likelihood estimator based on a consistent estimate of the spatial dependence parameter and sketch a proof of its consistency when the time dimension of the panel is fixed. In future versions of this paper, I plan to explore the small sample properties of the QML and constrained QML estimators with a Monte Carlo study. It would also be of interest to prove asymptotic normality of the proposed estimator as well as to derive the asymptotic properties of the QML estimator under some reasonable set of assumptions.

6. Appendix A - Derivatives of the QML Function

To judge whether the model is asymptotically identified, I inspect the Hessian of the concentrated likelihood function $Q_N(\theta)$. The first differential is

$$
\begin{aligned}
dQ_N &= -\frac{N}{2}\mathrm{tr}\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(d\boldsymbol{\Sigma}_{\Delta\eta}) && (55)\\
&\quad+\frac{N}{2}\mathrm{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\Delta\eta}^{-1}d\boldsymbol{\Sigma}_{\Delta\eta}\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N\right]\\
&\quad-\frac{N}{2}\mathrm{tr}\left[d\mathbf{R}'(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N\right]\\
&\quad-\frac{N}{2}\mathrm{tr}\left[\mathbf{R}'(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\Delta\eta}^{-1}(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W}')d\mathbf{R}\mathbf{S}_N\right]\\
&= \frac{N}{2}\mathrm{vec}\left[-\boldsymbol{\Sigma}_{\Delta\eta}'^{-1}+\boldsymbol{\Sigma}_{\Delta\eta}'^{-1}(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N'\mathbf{R}'(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\Delta\eta}'^{-1}\right]'\cdot\\
&\quad\cdot\mathbf{D}_{mT}\,d\mathrm{vech}\boldsymbol{\Sigma}_{\Delta\eta}^{-1}\\
&\quad N\mathrm{vec}\left[-\mathbf{S}_N'\mathbf{R}'(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\Delta\eta}'^{-1}(\mathbf{I}_{mN(T+1)}-\hat{\lambda}\mathbf{W}')\right]'d\mathrm{vec}\mathbf{R}
\end{aligned}
$$

where $\mathbf{D}_{mT}$ is a duplication matrix (such as that $\mathbf{D}_k\mathrm{vech}(\mathbf{X}) = \mathrm{vec}(\mathbf{X})$ for any $k \times k$ matrix $\mathbf{X}$), $\mathbf{K}_{sq}$ is a commutation matrix (such that $\mathbf{K}_{sq}\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(\mathbf{X}')$ for any $s \times q$ matrix $\mathbf{X}$),

$$
\begin{aligned}
d\mathrm{vech}\boldsymbol{\Sigma}_{\Delta\eta}^{-1} &= \mathrm{vech}\left[(\mathbf{A}_1 \otimes d\boldsymbol{\Psi}) + (\mathbf{A}_2 \otimes d\boldsymbol{\Omega})\right] && (56)\\
&= \mathbf{D}_{mT}^{-1}(I_T \otimes K_{m,T} \otimes I_m)(\mathrm{vec}\mathbf{A}_1 \otimes \mathbf{I}_{m^2})\mathbf{D}_{mT}\mathrm{vech}\boldsymbol{\Psi} + && (57)\\
&\quad+\mathbf{D}_{mT}^{-1}(I_T \otimes K_{m,T} \otimes I_m)(\mathrm{vec}\mathbf{A}_2 \otimes \mathbf{I}_{m^2})\mathbf{D}_{mT}\mathrm{vech}\boldsymbol{\Omega}\\
&= \mathbf{D}_{mT}^{-1}\mathbf{B}_1\mathrm{vech}\boldsymbol{\Psi} + \mathbf{D}_{mT}^{-1}\mathbf{B}_2\mathrm{vech}\boldsymbol{\Omega} && (58)
\end{aligned}
$$

and

$$
\begin{aligned}
d\mathrm{vec}\mathbf{R} &= \mathrm{vec}(\mathbf{A}_3 \otimes d\boldsymbol{\Phi}) && (59)\\
&= (I_T \otimes K_{m,T} \otimes I_m)(\mathrm{vec}\mathbf{A}_3 \otimes \mathbf{I}_{m^2})d\mathrm{vec}\boldsymbol{\Phi} && (60)\\
&= \mathbf{B}_3 d\mathrm{vec}\boldsymbol{\Phi} && (61)
\end{aligned}
$$

with $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ being matrices of constants reflecting the structure of $\boldsymbol{\Sigma}_{\Delta\eta}^{-1}$ and $\mathbf{R}$.

In particular,

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix} \tag{62}$$

$$\mathbf{A}_2 = \mathbf{I}_T - \mathbf{A}_1 \tag{63}$$

and

$$\mathbf{A}_3 = \begin{pmatrix} 0 & & & \cdots & 0 \\ -1 & 0 & & & \vdots \\ 0 & -1 & \ddots & & \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \cdots & 0 & -1 & 0 \end{pmatrix} \tag{64}$$

Defining

$$\mathbf{M}_1 = -\boldsymbol{\Sigma}_{\epsilon}'^{-1} + \boldsymbol{\Sigma}_{\epsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N'\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\epsilon}'^{-1} \tag{65}$$

and

$$\mathbf{M}_2 = -\mathbf{S}_N'\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\Delta\epsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}') \tag{66}$$

we can write the Jacobian of $Q_N(\theta)$ in a partitioned form as

$$DQ_N(\theta) = \frac{N}{2}\begin{bmatrix} \mathrm{vec}(\mathbf{M}_1)'\mathbf{B}_1 : \\ \mathrm{vec}(\mathbf{M}_1)'\mathbf{B}_2 : \\ 2\mathrm{vec}(\mathbf{M}_2)'\mathbf{B}_3 \end{bmatrix} \tag{67}$$

The second order differential is

$$
\begin{aligned}
d^2Q_N &= -\frac{N}{2}\operatorname{tr}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(d\boldsymbol{\Sigma}_{\varepsilon\varepsilon})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(d\boldsymbol{\Sigma}_{\varepsilon\varepsilon}) \\
&\quad + N\operatorname{tr}\left[\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}d\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}d\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N\right] \\
&\quad + 2N\operatorname{tr}\left[d\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}d\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N\right] \\
&\quad - N\operatorname{tr}\left[\mathbf{S}_N'\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}d\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Sigma}_{\varepsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')d\mathbf{R}\right] \\
&\quad - N\operatorname{tr}\left[\mathbf{S}_N'd\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')d\mathbf{R}\right] \\
&= -\frac{N}{2}(d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1})'\mathbf{D}_{mT}'(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon})\mathbf{D}_{mT}(d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}) \\
&\quad + N(d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon})'\mathbf{D}_{mT}'\left[\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N\mathbf{R}'(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\right] \cdot \\
&\quad \cdot\mathbf{D}_{mT}(d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}) \\
&\quad + 2N(d\operatorname{vec}\mathbf{R})\left[\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W}') \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N\right]\mathbf{D}_{mT}(d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}) \\
&\quad - N(d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon})'\mathbf{D}_{mT}'\left[\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\right](d\operatorname{vec}\mathbf{R}) \\
&\quad - N(d\operatorname{vec}\mathbf{R})'\left[\mathbf{S}_N \otimes (\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon\varepsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\right](d\operatorname{vec}\mathbf{R}) \\
&= \frac{N}{2}\begin{pmatrix}\mathbf{D}_{mT}\,d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \\ d\operatorname{vec}\mathbf{R}\end{pmatrix}'\mathbf{H}\begin{pmatrix}\mathbf{D}_{mT}\,d\operatorname{vech}\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \\ d\operatorname{vec}\mathbf{R}\end{pmatrix}
\end{aligned} \tag{68}
$$

where

$$\mathbf{H} = \begin{pmatrix}\mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{pmatrix} \tag{69}$$

$$\mathbf{H}_{11} = ([2\mathbf{M}_1 - \boldsymbol{\Sigma}_{\varepsilon\varepsilon}] \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon}) \tag{70}$$

$$\mathbf{H}_{12} = \left[\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\mathbf{R}\mathbf{S}_N \otimes \boldsymbol{\Sigma}_{\varepsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}')\right] \tag{71}$$

$$\mathbf{H}_{21} = \mathbf{S}_N \otimes (\mathbf{I}_{m(T+1)} - \hat{\lambda}\mathbf{W})\boldsymbol{\Sigma}_{\varepsilon}'^{-1}(\mathbf{I}_{mN(T+1)} - \hat{\lambda}\mathbf{W}') \tag{72}$$

$$\mathbf{H}_{22} = 2\mathbf{S}_N \otimes \mathbf{S}_N^{-1}\mathbf{M}_2 \tag{73}$$

Using our previous results, we have that

$$
\begin{aligned}
d^2Q_N &= \frac{N}{2}\begin{pmatrix}\mathbf{B}_1\mathrm{vech}\boldsymbol{\Psi}+\mathbf{B}_2\mathrm{vech}\boldsymbol{\Omega}\\ \mathbf{B}_3\mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}'\mathbf{H}\begin{pmatrix}\mathbf{B}_1\mathrm{vech}\boldsymbol{\Psi}+\mathbf{B}_2\mathrm{vech}\boldsymbol{\Omega}\\ \mathbf{B}_3\mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix} \qquad (74)\\
&= \frac{N}{2}\begin{pmatrix}\mathbf{B}_1\mathrm{vech}\boldsymbol{\Psi}\\ \mathbf{B}_2\mathrm{vech}\boldsymbol{\Omega}\\ \mathbf{B}_3\mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}'\begin{pmatrix}\mathbf{H}_{11} & \mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{11} & \mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{21} & \mathbf{H}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{B}_1\mathrm{vech}\boldsymbol{\Psi}\\ \mathbf{B}_2\mathrm{vech}\boldsymbol{\Omega}\\ \mathbf{B}_3\mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}\\
&= \frac{N}{2}\begin{pmatrix}\mathrm{vech}\boldsymbol{\Psi}\\ \mathrm{vech}\boldsymbol{\Omega}\\ \mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}'\begin{pmatrix}\mathbf{B}_1\mathbf{H}_{11}\mathbf{B}_1 & \mathbf{B}_1\mathbf{H}_{11}\mathbf{B}_2 & \mathbf{B}_1\mathbf{H}_{12}\mathbf{B}_3\\ \mathbf{B}_2\mathbf{H}_{11}\mathbf{B}_1 & \mathbf{B}_2\mathbf{H}_{11}\mathbf{B}_2 & \mathbf{B}_2\mathbf{H}_{12}\mathbf{B}_3\\ \mathbf{B}_3\mathbf{H}_{21}\mathbf{B}_1 & \mathbf{B}_3\mathbf{H}_{21}\mathbf{B}_2 & \mathbf{B}_3\mathbf{H}_{22}\mathbf{B}_3\end{pmatrix}\begin{pmatrix}\mathrm{vech}\boldsymbol{\Psi}\\ \mathrm{vech}\boldsymbol{\Omega}\\ \mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}\\
&= \frac{N}{2}\begin{pmatrix}\mathrm{vech}\boldsymbol{\Psi}\\ \mathrm{vech}\boldsymbol{\Omega}\\ \mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}'\mathbf{H}^*\begin{pmatrix}\mathrm{vech}\boldsymbol{\Psi}\\ \mathrm{vech}\boldsymbol{\Omega}\\ \mathrm{dvec}\boldsymbol{\Phi}\end{pmatrix}
\end{aligned}
$$

Hence the Hessian is then

$$\mathbf{H}Q_N = \frac{1}{2}\left[\mathbf{H}^* + \mathbf{H}^{*\prime}\right] \qquad (75)$$

7. Appendix B - Nuisance Property of λ

For the $m = 1$ case, the notation simplifies to:

$$\boldsymbol{\Sigma}_{\Delta\eta} = \begin{pmatrix}\psi & -\sigma^2 & & 0\\ -\sigma^2 & 2\sigma^2 & \ddots & \\ & \ddots & \ddots & -\sigma^2\\ 0 & & -\sigma^2 & 2\sigma^2\end{pmatrix} \qquad (76)$$

and $\mathrm{var}[\varepsilon_{i1} - (1-\phi)\xi_i] = \psi$ and $\mathrm{var}(\varepsilon_{it}) = \sigma^2$. If we assume that $\varepsilon_t \sim N(0; \sigma^2\mathbf{I}_N)$ and $[\varepsilon_1 - (1-\phi)\xi] \sim N(0; \psi\mathbf{I}_N)$ independent of ε_t for all $1 \leq t \leq T$, then the log-likelihood functions is:

$$
\begin{aligned}
\ln L_N(\theta|\Delta\mathbf{X}_t; \Delta\mathbf{y}_t; \Delta\mathbf{y}_{t-1}) &= -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}_{\Delta\eta}| + T\ln|\mathbf{I}_N - \lambda\mathbf{W}|\\
&\quad + \frac{1}{2}\mathbf{V}(\theta)'(\boldsymbol{\Sigma}_{\Delta\eta}^{-1}\otimes\mathbf{I}_N)\mathbf{V}(\theta) \qquad (77)
\end{aligned}
$$

where we define the sample counterpart of $\Delta\eta$ as

$$\mathbf{V}(\theta) = [\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda\mathbf{W})](\Delta\mathbf{Y} - \phi\Delta\mathbf{Y}_{-1}) \qquad (78)$$

with the vector of parameters $\theta = (\phi, \sigma^2, \psi, \lambda)'$. As before, we can express $\mathbf{\Sigma}_{\Delta\eta}$ as

$$\mathbf{\Sigma}_{\Delta\eta} = \begin{pmatrix} \psi & \sigma^2\mathbf{a}_1' \\ \sigma^2\mathbf{a}_1 & \sigma^2\mathbf{A}_2 \end{pmatrix} \tag{79}$$

where $\mathbf{a}_1 = (-1, 0, \ldots, 0)'$ and $\mathbf{A}_2$ are vector and a matrix of constants. The inverse of $\mathbf{\Sigma}_{\Delta\eta}$ is then

$$\mathbf{\Sigma}_{\Delta\eta}^{-1} = \frac{1}{d}\mathbf{A}^{-1} \tag{80}$$

where

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & -\mathbf{a}_1'\mathbf{A}_2^{-1} \\ -\mathbf{A}_2^{-1}\mathbf{a}_1 & \mathbf{I}_{T-1} - \mathbf{A}_2^{-1}\mathbf{a}_1\mathbf{a}_1'\mathbf{A}_2^{-1} \end{pmatrix} \tag{81}$$

and $d = \psi - \sigma^2\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1$. Using the same partitioning of $\mathbf{\Sigma}_{\Delta\eta}$, we can express its determinant as

$$|\mathbf{\Sigma}_{\Delta\eta}| = (\sigma^2)^{T-1}\psi|\mathbf{A}_2| + (\sigma^2)^T|\mathbf{A}_3| \tag{82}$$

where the $\mathbf{A}_3$ is equal to $\mathbf{A}_2$ with the first row replaced by $\mathbf{a}_1'$.

7.1. Partial Derivatives

The first and second-order partial derivatives of the likelihood function are:

$$\frac{\partial \ln L}{\partial \phi} = -\mathbf{V}'(\mathbf{\Sigma}_{\Delta\eta}^{-1} \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda\mathbf{W})]\Delta\mathbf{Y}_{-1} \tag{83}$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \psi} &= -\frac{N}{2|\mathbf{\Sigma}_{\Delta\eta}|}\left[(T-1)(\sigma^2)^{T-2}\psi|\mathbf{A}_2| + T(\sigma^2)^{T-1}|\mathbf{A}_3|\right] \\ &\quad + \frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\mathbf{V}'(\mathbf{A}^{-1}\otimes\mathbf{I}_N)\mathbf{V} \end{aligned} \tag{84}$$

where

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & -\mathbf{a}_1'\mathbf{A}_2^{-1} \\ -\mathbf{A}_2^{-1}\mathbf{a}_1 & \mathbf{I}_{T-1} - \mathbf{A}_2^{-1}\mathbf{a}_1\mathbf{a}_1'\mathbf{A}_2^{-1} \end{pmatrix} \tag{85}$$

and finally

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2|\mathbf{\Sigma}_{\Delta\eta}|}\left[(\sigma^2)^{T-1}|\mathbf{A}_2|\right] - \frac{1}{2d^2}\mathbf{V}'(\mathbf{A}^{-1}\otimes\mathbf{I}_N)\mathbf{V} \tag{86}$$

The second-order partial derivatives are:

$$\frac{\partial^2 \ln L}{\partial \phi \partial \lambda} = \mathbf{V}'(\boldsymbol{\Sigma}_{\Delta\varepsilon}^{-1} \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes \mathbf{W}]\Delta\mathbf{Y}_{-1} \tag{87}$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = \frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\mathbf{V}'(\mathbf{A}^{-1} \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes \mathbf{W}(\mathbf{I}_N - \lambda\mathbf{W})^{-1}]\mathbf{V} \tag{88}$$

$$\frac{\partial^2 \ln L}{\partial \psi \partial \lambda} = -\frac{1}{2d^2}\mathbf{V}'(\mathbf{A}^{-1} \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes \mathbf{W}(\mathbf{I}_N - \lambda\mathbf{W})^{-1}]\mathbf{V} \tag{89}$$

7.2. Probability Limits of the Hessian

The probability limits of the off-diagonal elements of the Hessian are:

7.2.1. $\partial\phi\partial\lambda$

$$p\lim_{N\to\infty} \frac{1}{N}\frac{\partial^2 \ln L}{\partial \phi \partial \lambda} = p\lim_{N\to\infty} \frac{1}{N}\Delta\varepsilon'(\boldsymbol{\Sigma}_{\Delta\varepsilon}^{-1} \otimes \mathbf{I}_N)(\mathbf{I}_T \otimes \mathbf{W})\Delta\mathbf{Y}_{-1} \tag{90}$$

The expected value of the first part of above expression is

$$E\frac{1}{N}\Delta\varepsilon'(\boldsymbol{\Sigma}_{\Delta\varepsilon}^{-1} \otimes \mathbf{I}_N)(\mathbf{I}_T \otimes \mathbf{W})\Delta\mathbf{Y}_{-1} = \frac{1}{N}\mathrm{tr}\left[E(\Delta\mathbf{Y}_{-1}\Delta\varepsilon')(\boldsymbol{\Sigma}_{\Delta\varepsilon}^{-1} \otimes \mathbf{I}_N)(\mathbf{I}_T \otimes \mathbf{W})\right] \tag{91}$$

To evaluate $E(\Delta\mathbf{Y}_{-1}\Delta\varepsilon')$, we first express $\Delta\mathbf{Y}_{-1}$ as a function of the model disturbances by recursive substitution:

$$\Delta\mathbf{Y}_{-1} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda\mathbf{W})]^{-1}\sum_{k=1}^{T-1}\phi^{k-1}\Delta\varepsilon_{-k} \tag{92}$$

and hence

$$E(\Delta\mathbf{Y}_{-1}\Delta\varepsilon') = [\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda\mathbf{W})]^{-1}\sum_{k=1}^{T-1}\phi^{k-1}E(\Delta\varepsilon_{-k}\Delta\varepsilon') \tag{93}$$

Now, $E(\Delta\varepsilon_{-k}\Delta\varepsilon')$ has a structure similar to $(\boldsymbol{\Sigma}_{\Delta\varepsilon} \otimes \mathbf{I}_N)$:

$$E(\Delta\varepsilon_{-k}\Delta\varepsilon') = (\boldsymbol{\Sigma}_{\Delta\varepsilon_{-k}} \otimes \mathbf{I}_N) \tag{94}$$

where the first kN rows of $\mathbf{\Sigma}_{\Delta\varepsilon_{-k}}$ are zeros and the remaining $(T-k)N$ rows are the first $(T-k)N$ rows of $\mathbf{\Sigma}_{\Delta\varepsilon}$. Therefore, we have

$$\begin{aligned} E(\Delta\varepsilon_{-k}\Delta\varepsilon')(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N) &= (\mathbf{\Sigma}_{\Delta\varepsilon_{-k}}\otimes\mathbf{I}_N)(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N) \\ &= (\mathbf{I}_{T_{-k}}\otimes\mathbf{I}_N) \end{aligned} \tag{95}$$

where the first k rows of $\mathbf{I}_{T_{-k}}$ are zeros and remaining $(T-k)$ rows are first $(T-k)$ rows of a $T\times T$ identity matrix. Collecting the results, we have that

$$\begin{aligned} E(\Delta\mathbf{Y}_{-1}\Delta\varepsilon')(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N) &= [\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}\sum_{k=1}^{T-1}\phi^{k-1}E(\Delta\varepsilon_{-k}\Delta\varepsilon')(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N) \\ &= [\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}\sum_{k=1}^{T-1}\phi^{k-1}(\mathbf{\Sigma}_{\Delta\varepsilon_{-k}}\otimes\mathbf{I}_N)(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N) \\ &= [\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}\sum_{k=1}^{T-1}\phi^{k-1}(\mathbf{I}_{T_{-k}}\otimes\mathbf{I}_N) \\ &= [\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}(\mathbf{\Phi}\otimes\mathbf{I}_N) \end{aligned} \tag{96}$$

where $\mathbf{\Phi}$ is a matrix of zeros except for $\frac{T(T-1)}{2}$ elements below the main diagonal which are powers of ϕ. Therefore,

$$\begin{aligned} \operatorname{tr}\left[E(\Delta\mathbf{Y}_{-1}\Delta\varepsilon')(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W})\right] &= \operatorname{tr}\left[[\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}(\mathbf{\Phi}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W})\right] \\ &= \operatorname{tr}\left[(\mathbf{I}_T\otimes\mathbf{W})[\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}(\mathbf{\Phi}\otimes\mathbf{I}_N)\right] \\ &= \operatorname{tr}\left[(\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1})(\mathbf{\Phi}\otimes\mathbf{I}_N)\right] \\ &= \operatorname{tr}\left[(\mathbf{I}_T\mathbf{\Phi}\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1})\right] \\ &= T\cdot\operatorname{tr}\left[\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1})\right] \\ &< T\cdot\varpi \end{aligned} \tag{97}$$

where ϖ is such that $\forall N:\forall 0\le i\le N:\sum_{i=1}^{N}|a_{ij}|<\varpi$ where a_{ij} are elements of $\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1}$. Existence of such ϖ is due to absolute summability of $\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1}$. Hence we can conclude that

$$E\frac{1}{N}\Delta\varepsilon'(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W})\Delta\mathbf{Y}_{-1}<\frac{T}{N}\varpi\to 0 \tag{98}$$

The limit of the variance of the same expression is

$$\operatorname*{plim}_{N\to\infty}\operatorname{var}\frac{1}{N}\Delta\varepsilon'(\mathbf{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W})\Delta\mathbf{Y}_{-1}$$

$$
\begin{aligned}
&= \operatorname*{p\,lim}_{N\to\infty} \frac{1}{N^2}\Delta\mathbf{Y}'_{-1}(\mathbf{I}_T\otimes\mathbf{W})'(\boldsymbol{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W})\Delta\mathbf{Y}_{-1} \\
&= \operatorname{tr}\left[\operatorname*{p\,lim}_{N\to\infty} \frac{1}{N^2}\Delta\mathbf{Y}_{-1}\Delta\mathbf{Y}'_{-1}(\mathbf{I}_T\otimes\mathbf{W})'(\boldsymbol{\Sigma}_{\Delta\varepsilon}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W})\right] \\
&= 0 \qquad (99)
\end{aligned}
$$

by assumption (2) below. Therefore, by Chebychev Lemma, the probability limit of $\frac{1}{N}\frac{\partial^2}{\partial\rho\partial\lambda}$ is zero.

7.2.2. $\partial\sigma^2\partial\lambda$

For the term $\frac{\partial^2 \ln L}{\partial\sigma^2\partial\lambda}$ we have

$$
\operatorname*{p\,lim}_{N\to\infty}\frac{1}{N}\frac{\partial^2 \ln L}{\partial\sigma^2\partial\lambda} = \frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\operatorname*{p\,lim}_{N\to\infty}\frac{1}{N}\Delta\varepsilon'(\mathbf{A}^{-1}\otimes\mathbf{I}_N)[\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1}]\Delta\varepsilon \qquad (100)
$$

To get the needed result, we again show that expected value and variance of the above expression is zero in the limit, and hence by Chebyshev Lemma its probability limit is zero:

$$
\begin{aligned}
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\sigma^2\partial\lambda} &= \frac{1}{N}\frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}E\operatorname{tr}[\frac{1}{N}(\Delta\varepsilon\Delta\varepsilon')(\mathbf{A}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1})] \\
&= \frac{1}{N}\frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\operatorname{tr}[E(\Delta\varepsilon\Delta\varepsilon')(\mathbf{A}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1})] \\
&= \frac{1}{N}\frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\operatorname{tr}[(d\mathbf{A}\otimes\mathbf{I}_N)(\mathbf{A}^{-1}\otimes\mathbf{I}_N)(\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1})] \\
&= \frac{1}{N}\frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d}\operatorname{tr}[\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1}]\to 0 \qquad (101)
\end{aligned}
$$

since both $\mathbf{A}^{-1}$ and $\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1}$ are absolutely summable. The variance is

$$
\begin{aligned}
\operatorname{var}\frac{1}{N}\frac{\partial^2 \ln L}{\partial\sigma^2\partial\lambda} &= \operatorname{var}\frac{1}{N}\frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\operatorname{tr}[(\Delta\varepsilon\Delta\varepsilon')(\mathbf{A}^{-1}\otimes\mathbf{I}_N)[\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}] \\
&= \frac{1}{N^2}\left(\frac{\mathbf{a}_1'\mathbf{A}_2^{-1}\mathbf{a}_1}{2d^2}\right)^2 \cdot \qquad (102) \\
&\quad\cdot\operatorname{tr}\left\{\begin{array}{c}(\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W}')]^{-1}\mathbf{W}')(\mathbf{A}'^{-1}\otimes\mathbf{I}_N)\operatorname{var}(\Delta\varepsilon\Delta\varepsilon')\\ \cdot(\mathbf{A}^{-1}\otimes\mathbf{I}_N)[\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}\end{array}\right\}
\end{aligned}
$$

If $\Delta\varepsilon_t$ has finite fourth moments, we have

$$
\begin{aligned}
\operatorname{var}\frac{1}{N}\frac{\partial^2 \ln L}{\partial\sigma^2\partial\lambda} &< \text{const}\cdot\frac{1}{N^2}\operatorname{tr}[(\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W}')^{-1}]\mathbf{W}') \\
&\quad\cdot(\mathbf{A}'^{-1}\otimes\mathbf{I}_N)(\mathbf{A}^{-1}\otimes\mathbf{I}_N)[\mathbf{I}_T\otimes\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})]^{-1}] \\
&= \text{const}\cdot\frac{1}{N^2}\operatorname{tr}[(\mathbf{I}_T\otimes(\mathbf{I}_N-\lambda\mathbf{W}')^{-1}] \\
&\quad\cdot\mathbf{W}'\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})]^{-1})(\mathbf{A}'^{-1}\mathbf{A}^{-1}\otimes\mathbf{I}_N) \\
&\to 0 \qquad (103)
\end{aligned}
$$

because $(\mathbf{I}_N-\lambda\mathbf{W}')^{-1}\mathbf{W}'\mathbf{W}(\mathbf{I}_N-\lambda\mathbf{W})^{-1}$ and $\mathbf{A}'^{-1}\mathbf{A}^{-1}$ are absolutely summable.

8. References

Ahn, S.C. and P. Schmidt (1995): Efficient Estimation of Models for Dynamic Panel Data, Journal of Econometrics, 68, 5-27.

Anselin, L. (1988): Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Boston.

Arellano, M. and S.R. Bond (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and Application to Employment Equations, Review of Economic Studies, 58, 277-297.

Arellano, M. and O. Bover (1995): Another Look at the Instrumental Variable Estimation if Error-Component Models, Journal of Econometrics, 68, 28-51.

Binder, Michael, Cheng Hsiao and M. Hashem Pesaran (2001): Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, mimeo.

Binder, Michael, Cheng Hsiao, Jan Mutl and M. Hashem Pesaran (2002): Computational Issues in Estimation of Higher-Order Panel Vector Autoregressions, mimeo.

Chang, Yoosoon (2001): Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, mimeo.

Chen, Xiaohong and Timothy Conley (2000): A New Semiparametric Spatial Model for Panel Time Series, mimeo.

Cliff, A. and J. Ord (1973): Spatial Autocorrelation, Pion, London.

Cliff, A. and J. Ord (1981): Spatial Processes, Models and Applications, Pion, London.

Hsiao, C., M. H. Pesaran and A. K. Tahmiscioglu (2001): Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, Journal of Econometrics, forthcoming.

Kelejian, Harry H. and Ingmar R. Prucha (1998): A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, Journal of Real Estate Finance and Economics, 17, 99-121.

Kelejian, Harry H. and Ingmar R. Prucha (1999): A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, International Economic Review, 40, 509-533.

Kelejian, Harry H. and Ingmar R. Prucha (2001): Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations, mimeo.

Kelejian, Harry H. and Ingmar R. Prucha (2001): On the Asymptotic Distribution of the Moran I Test Statistic with Applications, Journal of Econometrics, 104, 219-257..

O'Connel, paul G. J. (1998): The Overvaluation of Purchasing Power Parity, Journal of International Economics, 44, 1-19.

Pinkse, Joris (1999): Asymptotic Properties of Moran and Related Tests and Testing for Spatial Correlation in Probit Models, mimeo.

Whittle, P. (1954): On Stationary Processes in the Plane, Biometrica, 41, 434-449.